多階層記憶におけるデータ並べかえと記憶階層の最適化†

佐藤隆士†† 津田孝夫†††

現在の大型計算機では，高速小容量から低速大容量に至る多くの階層からなる記憶装置をもっている．本論文では，筆者が以前発表した2階層記憶でのデータ任意並べかえアルゴリズム $d\log_w d$ 法を多階層記憶にも適用できるようにした拡張 $d\log_w d$ 法を提案する．多階層記憶をモデル化し，アルゴリズムのコストを計算することにより，限定されたアルゴリズムに対してではあるが多階層記憶の各レベルを統一的に扱うことを試みた．その結果，1) 2階層記憶の中間に，アクセススピードに見合った容量をもつ記憶階層を設けると，提案の方法により，2階層記憶にくらべ“レベル2のアクセススピード／最下位レベルのアクセススピード”に比例して高速にデータの並べかえができることがわかった．また，2) 中間レベルに必要十分な記憶容量，3) 性能を変えずに中間の隣合った2つのレベルを1つのレベルでおきかえるとき，そのレベルの特性と容量，を求めることができた．

1. まえがき

現在の大型計算機では，高速小容量から低速大容量の順に，キャッシュ，主記憶，ディスクキャッシュ，磁気ディスク，MSS 等，多くの階層からなる記憶装置をもっている．これらに加え，最近では磁気バブル，光ディスク等の記憶装置が登場しているが，今後とも，計算機の記憶は，ますます多レベル化，多様化することが予想される．

従来，一般的なプログラムの振舞いに対して，各レベルの記憶装置の特性にあった記憶管理方式が検討されてきた．しかし筆者は，システムとして記憶階層を効率的に使用する手法の確立が要請されていると考える．本論文では，多階層記憶をモデル化し，限定されたアルゴリズムに対して，統一的に扱うことを試みた．

筆者らは，2階層記憶でのデータ並べかえについて研究してきたが[1)~3)]，本論文では，2階層記憶での任意並べかえアルゴリズム $d\log_w d$ 法[3)]を多階層記憶に拡張した拡張 $d\log_w d$ 法を提案する．モデル化された多階層記憶について解析をした結果，1) 2階層記憶の中間に，アクセススピードに見合った容量をもつ記憶階層を設けると，提案の方法により，2階層記憶にくらべ“レベル2のアクセススピード／最下位レベルのアクセススピード”に比例して高速にデータの並べかえができることがわかった．また，2) 中間レベルに必要十分な記憶容量，3) 性能を変えずに中間の隣合った2つのレベルを1つのレベルでおきかえるとき，そのレベルの特性と容量，を求めることができた．

2. 定　義

記憶は n 階層とし，最下位レベル（レベル n）のデータを並べかえの対象とする．レベル i ($i=1,2,\cdots,n-1$) のメモリはレベル $i+1$ のメモリと“i 次ページ”と呼ぶ転送単位で，データを移動できる．レベル $i+1$ から i 方向の転送を読込み，逆方向の転送を書出しという．両方向の転送回数は等しくなるので，本論文では，読込み回数のみ勘定する．また，この読込みを i 次フェッチ（次数がわかる場合は単にフェッチ）ともいう．データ並べかえに要するコストは，各レベル間の読込み回数とする．レベル i ($i=1,2,\cdots,n$) のメモリ量は i 次ページを単位として w_i ページ分とする．ここまでの定義を図1に示す．

i ($i=1,2,\cdots,n$) 次ページの1ページに $i-1$ 次ページがちょうど p_i 格納可能であるとする．0次ページは，データの単位であり，レコードという．例えばレベル2が格納可能な w_2 2次ページは，1次ページ数で w_2p_2，0次ページすなわちレコード数でいうと $w_2p_2p_1$ である．i ($i=1,2,\cdots,n$) レベルの記憶を j ($j=0,1,\cdots,i$) 次ページを単位に呼ぶとき，ページ番号を使用する．ページ番号は論理的に番地の小さい場所から順に付けられ0から $w_i\prod_{k=j+1}^{i}p_k-1$ までである（ただし，$i=j$ のときは0から w_i-1）．並べかえ対象となるのは，レベル n に格納された p_n' 個の $n-1$ 次ページ分のデータであるとする．レベル n に格納され

† Permuting Information in Multi-Level Storage and the Optimization of Memory Hierarchy by TAKASHI SATO (Department of Information Engineering, Takuma National College of Technology) and TAKAO TSUDA (Department of Information Science, Kyoto University).

†† 鈍間電波工業高等専門学校情報工学科
††† 京都大学工学部情報工学科

ているので，$p_n' \leqq w_n p_n$ である．これは $p_n' \prod_{j=1}^{n-1} p_j$ レコードである．また，レベル n の必要性から，$p_n' > w_{n-1}$ でなければならない．

3. 3階層記憶

本章では，多階層記憶での議論の理解を容易にするため，$n=3$ すなわち3階層記憶の場合について述べる．**図2**は図1を $n=3$ について書き直したものである．

3.1 レベル2，3間のデータ転送

レベル3の p_3' 2次ページに格納されている $p_3'p_2$ 1次ページをレベル2の w_2 2次ページで並べかえる．まず，簡単のため p_3' が w_2 のべき乗の場合について述べ，後で一般の場合に拡張できることを示す．アルゴリズムは，$d\log_w d$ 法[3]を多階層記憶用にした拡張 $d\log_w d$ 法である．

レベル2，3間の転送では，レベル3上の2次ページ間で1次ページを並べかえることが目的である．

並べかえは，$t \triangleq \log_{w_2} p_3'$ の段階からなる．これを(レベル2，3間の)第1から第 t パスということにする．

[定義] 第 j パス $(j=0,1,2,\cdots,t)$ のグループ u $(u=0,1,\cdots,w_2{}^j-1)$ とは，最終状態（目標状態）のレコード並びで，$\{u, w_2{}^j+u, 2\cdot w_2{}^j+u, \cdots, p_3'-w_2{}^j+u\}$ の番号の2次ページに含まれるレコードのみからできている2次ページのことである（文献3）参照）．□

第 i パスでは第 $i-1$ パスのグループ u を第 i パスのグループ u，グループ $w_2{}^{i-1}+u, \cdots$，グループ $(w_2-1)\cdot w_2{}^{i-1}+u$ に分割する．ここで $i=1,2,\cdots,t$，第0パスのグループは1つだけでその要素は並べかえ対象となる全2次ページである．

次に，分割の方法を述べる．第 $i-1$ パスのグループ u をレベル3からレベル2に順次フェッチする．フェッチされた各2次ページはレベル1を用い，3.2節で述べる方法で，分解再編成され，第 i パスのグループに組み立てられる．レベル2に全部2次ページが読み込まれると，w_2 個の2次ページが収まるが，これにより少なくとも1つの第 i パスの2次ページができる（文献3）の補題参照）ので，それをレベル3に追い出す．それがグループ $j\cdot w_2{}^{i-1}+u$ $(j=0,1,\cdots,w_2-1)$ に属するとき，次のパスではこのグループの番号，すなわち $j\cdot w_2{}^{i-1}+u$ から始まる $w_2{}^i$ 間隔のページ番号で参照できるようページの呼びかえを行っておく．

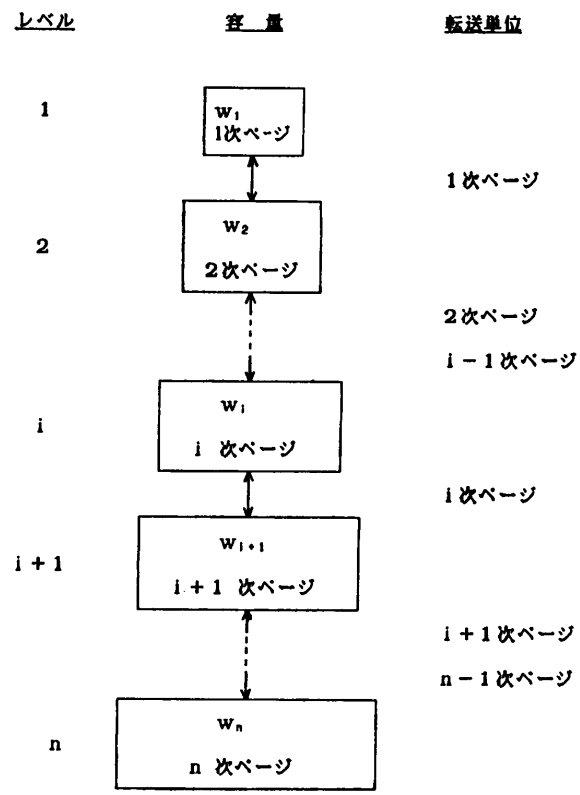

図1 記憶階層の概念図
Fig. 1 Notion of multi-level storage.

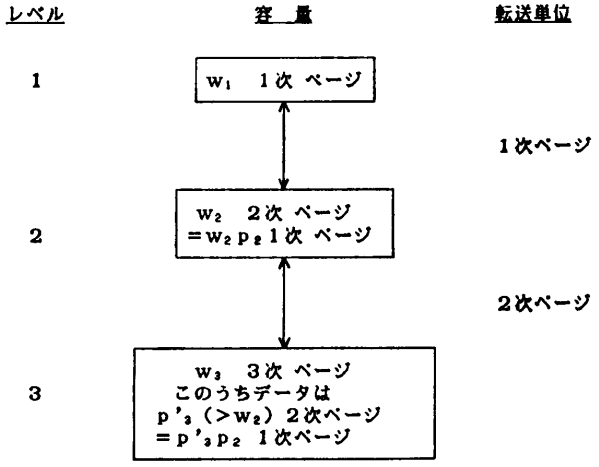

図2 3階層記憶の概念図
Fig. 2 Notion of 3-level storage.

この操作を第1から第 t パスまでのすべてのグループ u について行うと，レベル3の1次ページの並びは最終状態になる．レベル3から2への2次フェッチ数 F_2 は，

$$F_2 \leqq p_3' \log_{w_2} p_3' \quad \text{2次フェッチ} \qquad (1)$$

である．不等号の意味は，あるパスにおいて，次のパ

スの2次ページにもなっているページはフェッチ不要であることによる．

p_3' が w_2 のべき乗でないときでも，2次ページの番号が $p_3'-1$ 以下であることに注意すれば，同じアルゴリズムが使用できる．このとき，フェッチ数は，

$$F_2 \leqq p_3' \lceil \log_{w_2} p_3' \rceil \quad 2\text{次フェッチ} \qquad (2)$$

である．

2次ページ番号		2次ページ番号		2次ページ番号	
0	0 1 2 3	0	0 0 0 0	0	0 2 0 2
1	0 1 2 3	1	1 1 1 1	1	1 3 1 3
2	0 1 2 3	2	2 2 2 2	2	0 2 0 2
3	0 1 2 3	3	3 3 3 3	3	1 3 1 3
	(a) 初期状態		(b) 最終状態		(c) パス1終了時

図3 3階層記憶でのレコード並べかえ例—その1—
Fig. 3 Example of rearranging records on 3-level storage—1.

3.2 レベル1，2間のデータ転送

A) $w_1 \geqq w_2$ のとき

レベル2，3間の第 i パスではレベル2の w_2 2次ページに第 $i-1$ パスのグループ u ($u=0,1,\cdots,w_2^{i-1}-1$) が読み込まれているので，これを前節で述べたように w_2 個のグループに分割しなければならない．この作業は，各レコード（0次ページ）がグループ別に w_2 種類あるので，それを種類ごとに集めることであるとも言える．そのとき必要となる1次ページ内のレコード並べかえは CPU とレベル1間で行われる．

レベル3からレベル2に読み込まれた第 $i-1$ パスのグループ u に属する各2次ページは p_2 個の1次ページからなるが，それを順次レベル1に読み込む．レベル1に w_2 ($\leqq w_1$) 1次ページまで読み込まれるとレコードは w_2 種類だから，少なくとも1つ第 i パスのグループを集めた1次ページができる．それをレベル2に追い出し，次の（まだレベル1に読み込まれていない）1次ページをレベル2から読み込む．次の1次ページがないときは，レベル2には少なくとも1つ第 i パスの2次ページができているので，それをレベル3に追い出し，レベル3からグループ u の新しい2次ページを読み込んで以上の操作を繰り返す．新しい2次ページがないときはレベル2上の2次ページはすべて第 i パスのページになっているので，レベル3に追い出し，グループ u について終了する．第 i パスは，この操作を第 $i-1$ パスのグループ u すべてについて行う．レベル1の w_1-w_2 個の1次ページは使用されない．

レベル2から1への1次フェッチ数について考える．1回2次フェッチあたり p_2 1次フェッチが必要だから F_2p_2 となる．しかし，このままでは，2次ページ内での1次ページの順は希望どおりになるとはかぎらない．これを保証するためには，最終パスにおいてできたレベル3上の2次ページを並べかえる必要がある．この作業をレベル1の w_1 のうち w_2 1次ページだけを使って $d\log_w d$ 法で行うと，1つの2次ページあたり $p_2\log_{w_1}p_2$ 1次フェッチ以下でできるはずである．p_3' 2次ページ全部で，$p_2p_3'\log_{w_1}p_2$ 1次フェッチ以下である．したがって，レベル1，2間の1次フェッチ数 F_1 は，

$$F_1 \leqq p_2F_2 + p_2p_3'\log_{w_1}p_2 \leqq p_2p_3'\log_{w_1}p_2p_3' \qquad (3)$$

となる．p_2, p_3' が w_2 のべき乗でない時は対数に天井（$\lceil\ \rceil$）が付くはずであるが，近似値として天井をはずしておいた．なお，上式で，p_2p_3' はレベル3のデータ量を1次ページで数えたページ数になっている．

結局，レベル1の w_1 1次ページのうち w_1-w_2 は使用せずに式(3)を得たので，$w_1=w_2$ で十分であるといえる．

[例1] 各レコードを最終状態で行くべきレベル3の2次ページ番号で表し，レベル3の初期状態を**図3**(a)，最終状態を同図(b)とする．$p_3'=4$, $p_1=p_2=2$, $w_1=w_2=2$ とする．4×4 行列データの転置例である．

レベル2のパス数は，$t=\log_{w_2}p_3'=2$ である．第1パス終了時には，図3(c)に示すようにする．第2パス終了時には，もちろん図3(b)のようにならなければならない．第1パスの詳細を**図4**に，第2パスの詳細を**図5**に示す．図中，矢印はデータの移動（コピー）を示す．レベル1内ではレコード単位，レベル1，2間では1次ページ単位，レベル2，3間では2次ページ単位である．移動の順番は，(a)(b)(c)…の順，さらに同じアルファベットが付けられた図の中では，矢印に付けられた番号順である．四角は各レベルのページを表している．縦線で区切られた領域は上位レベルとの転送単位である．レベル3は書かれていないので，図3を参考にしていただきたい．

たとえば，図4では次のようになる．まず(a)でレベル3のページ0を読み込むと(b)のようになる．○印が付けられているのは，この1次ページ分2レコー

ドは，これからレベル1に読み込まれて処理されるべきであることを示している．次に(b)でレベル2の0ページのうち，“0”，“1”で表された2レコードをレベル1のページ0に読み込むと(c)のようになる．×印が付けられているのは，この1次ページ分のレコードは不要であり，書きかえられてもかまわないことを示している．(c)では“2”，“3”のレコードをレベル1のページ1に読み込む．(d)では①でレコードを交換し，②でレベル1から2へ書き出し，③でレベル3から2へ読込みを行うと(e)のようになる．●印が付けられているのは，この1次ページはレベル1から2へ書き出されたばかりでまだレベル3にまで書き出されていないので，書きかえができないことを示している．図に書かれたレコードの並びは，1つ前のアルファベットが付けられた図で矢印の操作を完了した時点の状態である．以下，同様にして(j)までの操作で第1パスを終了する．

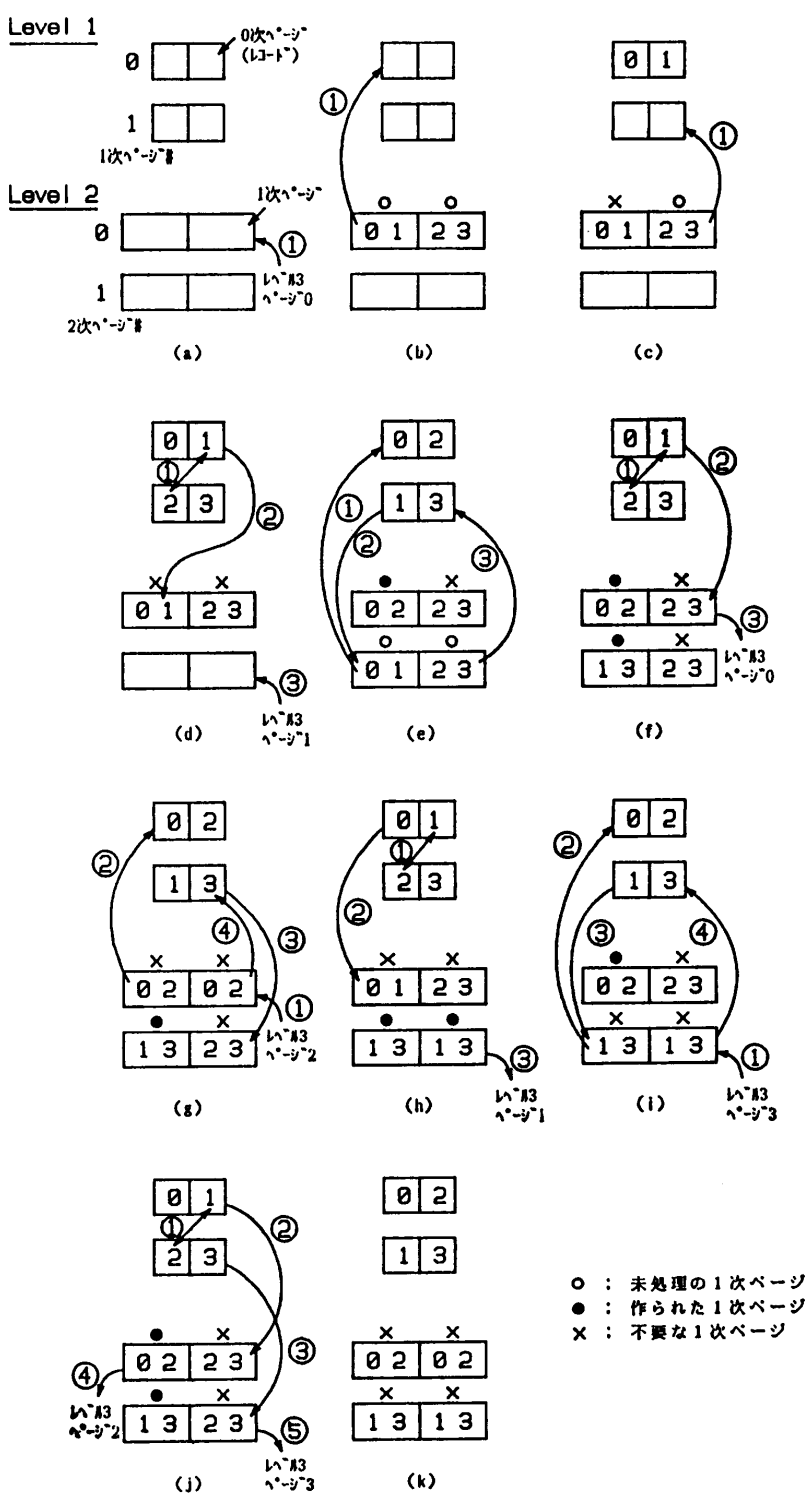

図4　第1パスのレコード移動の詳細
Fig. 4　Details of rearranging records in the first pass.

アルゴリズムは前節と本節で述べたとおりである．図では読込みを中心に見ていくとわかりやすい．書出しは，読込みに必要なスペースを確保するために読込みの直前に行われると考えればよい．なお，この例では最終パスでできた2次ページを並べかえる $p_2 \log_{w_2} p_2$ フェッチ以下は0（フェッチ不要）になる．　□

B)　$w_1 < w_2$ のとき

レベル3から2に読み込まれた1つの2次ページは p_2 個の1次ページからなっている．レベル2, 3間の並べかえ要求により，w_2 種類のレコードを w_1 ($<w_2$) 個の1次ページで分類する必要がある．このため，A）の場合のように p_2 回の1次フェッチ，すなわちひと通りレベル1に読み込むだけでは無理である．レベル1，2間についてもレ

ベル 2, 3 間と同様な方法により p_2 1次ページを $t_2=\lceil \log_{w_1} w_2 \rceil$ パスで処理する．すなわち，レベル 1, 2 間の第 i パス ($i=1,2,\cdots, t_2$) で，0 から w_2-1 のページ番号を w_1 進法で表したとき下 i 桁までについてレコードがそろうようにする（例2参照）．

転送回数は1回の2次フェッチあたり $p_2 t_2 = p_2 \lceil \log_{w_1} w_2 \rceil$ 1次フェッチである．したがって，A) で述べたレベル 2, 3 間の最終パス終了後の並べかえも考慮すると，1次フェッチ数の総計 F_1 は，

$$F_1 \leqq p_2 \lceil \log_{w_1} w_2 \rceil F_2 + p_2 p_3' \lceil \log_{w_1} p_2 \rceil, \quad (4)$$

近似的には「 」をはずして，

$$F_1 \leqq p_2 p_3' \log_{w_1} p_2 p_3' \quad (5)$$

となる．

[例2] レベル3の初期状態を**図6**(a)，最終状態を同図(b)とする．$p_3'=4$, $p_1=3$, $p_2=2$, $w_1=2$, $w_2=4$ のときの並べかえ例である．レベル 2, 3 間のパス数は，$t=\log_{w_2} p_3'=1$, レベル 1, 2 間パス数は $t_2=\lceil \log_{w_1} w_2 \rceil = 2$ となる．**図7**(a)～(g)にレコード並べかえの詳細を示す． □

C) w_1, w_2 の決め方

式(3)，(5)から $w_1 \leqq w_2$ で，1次フェッチ数のめやすは，

$$F_1 \leqq p_2 p_3' \log_{w_1} p_2 p_3' \quad (6)$$

である．式(1)あるいは(2)から，F_2 を小さくするためにはレベル2の2次ページ数 w_2 を大きくしたほうが，また式(6)から，F_1 を小さくするためには，w_1 を大きくしたほうがよいように見える．しかし，現実に制限となるのは w_1, w_2 の値よりむしろ各レベルの記憶容量である．そこで，レベル 1, 2 の記憶容量と，レベル3のデータ量をレコード数で，

$$K_1=p_1 w_1,\ K_2=p_1 p_2 w_2,\ K_3=p_1 p_2 p_3' \quad (7)$$

と表す．式(2)も(1)で近似し，式(1)と(6)に式(7)を代入すると，

$$F_2 \leqq (K_3/K_2) w_2 \log_{w_2} (K_3/K_2) w_2 \quad (1)'$$

$$F_1 \leqq (K_3/K_1) w_1 \log_{w_1} (K_3/K_1) w_1 \quad (6)'$$

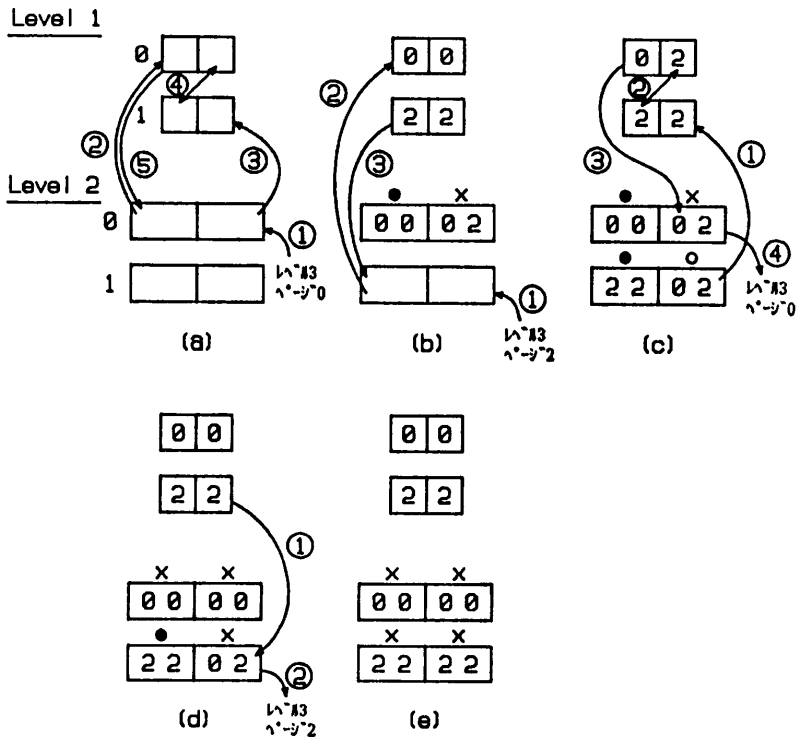

図5 第2パスのレコード移動の詳細（第1パスのグループ0について）
Fig. 5 Details of rearranging records in the second pass.

2次ページ番号						
0	0	1	1	1	2	2
1	1	1	1	3	3	3
2	0	0	0	2	3	3
3	0	0	2	2	2	3

(a) レベル3 初期状態

2次ページ番号						
0	0	0	0	0	0	0
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3

(b) レベル3 最終状態

図6 3階層記憶でのレコード並べかえ例—その2—
Fig. 6 Example of rearranging records on 3-level storage—2.

となる．式(1)′の右辺を $f_2(w_2)$ とおき w_2 で微分する．

$$\frac{d f_2(w_2)}{d w_2} = \frac{K_3}{K_2}\left\{\frac{\ln(K_3/K_2)}{\ln w_2}\left(1-\frac{1}{\ln w_2}\right)+1\right\} \quad (8)$$

例1では説明を短くするために $w_2=2$ としたが，現実的には $w_2>e$ なので，$f_2(w_2)$ は w_2 の増加関数であり，w_2 は小さいほうが2次フェッチ数を減少することができる．同様に式(6)′から，w_1 ($\leqq w_2$) も小さいほうが1次フェッチ数を減少できることがわかる．

現実的には，各レベルの記憶容量を一定にして，w_1, w_2 をいくらでも小さくすることはできない．なぜなら，拡張 $d \log_w d$ 法による場合，各レベルの転

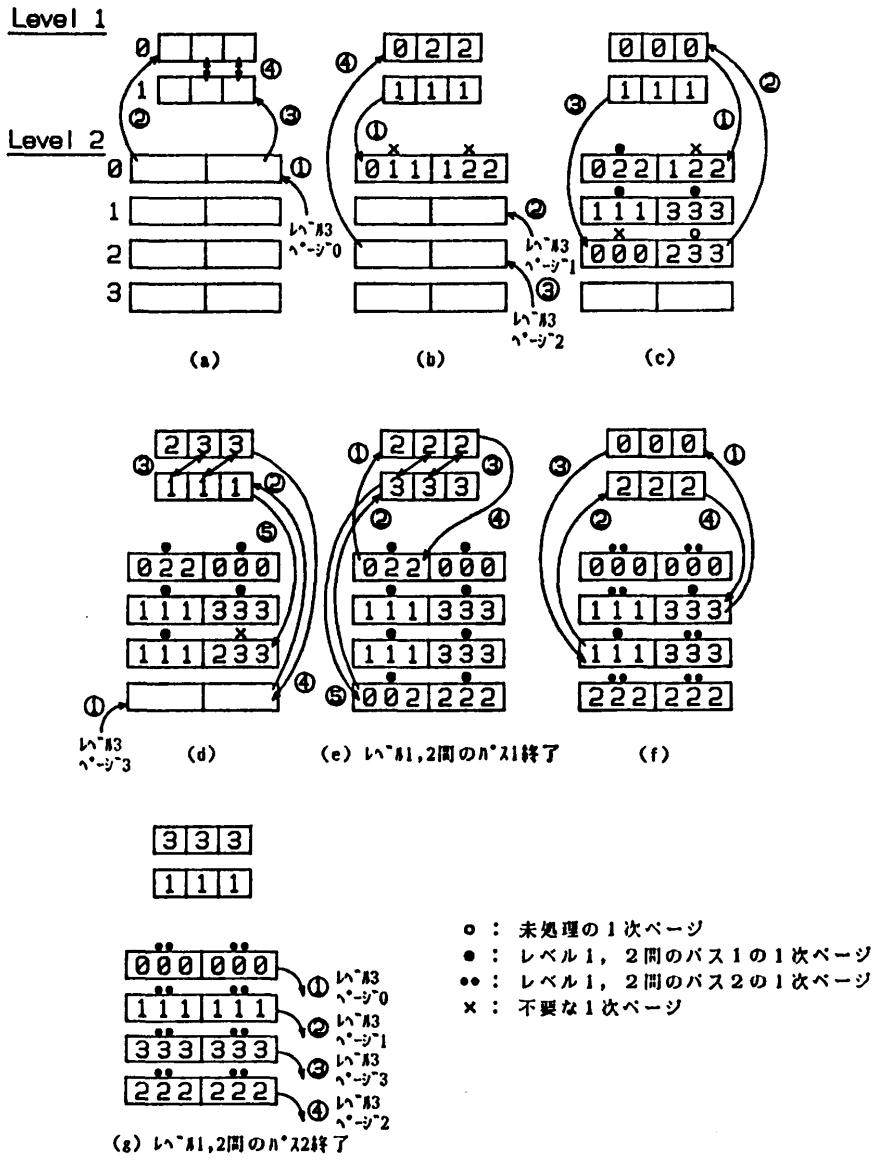

図 7 例2のレコード並べかえの詳細
Fig. 7 Details of rearranging records of the example 2.

送単位 p_1 あるいは p_1p_2 レコードは，記憶装置の特性により，上限が決まっているので式(7)から w_1, w_2 に下限が存在するからである．

4. n 階層記憶の場合

n 階層記憶の場合には，レベル n にある p_n' 個の $n-1$ 次ページ分のレコードを並べかえることになる．3章と同じ理由で，各レベルの記憶領域を有効に利用するためには，$w_1 \leqq w_2 \leqq \cdots \leqq w_{n-1}$ でなければならない．このとき，レベル $n-1, n$ 間のパス数は $\log_{w_{n-1}} p_n'$，レベル $i-1, i$ 間 $(i=2,3,\cdots,n-1)$ のパス数は，$\log_{w_{i-1}} w_i$ となる．ただし，パス数は整数だが，近似的に⌈ ⌉をはずしてある．以降の解析でも暗黙のうちにこの近似を行っている．

フェッチ数については $d \log_w d$ 法を用いると，レベル $n-1, n$ 間で，

$$F_{n-1} \leqq p_n' \log_{w_{n-1}} p_n' \quad n-1 \text{ 次フェッチ，} \quad (9)$$

レベル $i-1, i$ 間 $(i=2,3,\cdots,n-1)$ は，レベル $i-1$ の w_{i-1} 個の $i-1$ 次ページを使ってレベル i の w_i 種類のレコードを分類する．1つの i 次ページに付き $p_i \log_{w_{i-1}} w_i$，$i-1$ 次フェッチだから，

$$F_{i-1} \leqq p_i \log_{w_{i-1}} F_i + q_i \log_{w_{i-1}} p_i$$

$i-1$ 次フェッチ (10)

$$q_i = p_n' \prod_{j=i}^{n-1} p_j \tag{11}$$

となる．ここで，q_i は全レコードを $i-1$ 次ページで数えたページ数である．$q_i \log_{w_{i-1}} p_i$ の項は最終パス終了後のレベル i の i 次ページを並べかえるのに必要なフェッチ数である．

式(10)の漸化式を式(9)を初期値として解くと，

$$F_i \leqq q_{i+1} \log_{w_i} q_{i+1}$$

i 次フェッチ $(i=1, 2, \cdots, n-1)$ (12)

となる．

5. 記憶階層下での転送コストについて

前章までで求めたフェッチ数から5.1節では記憶階層下での全転送コストを定義し，それをもとに5.2節では2，3，4階層記憶の比較を行う．5.3節では一般化し，n (≧3) 階層記憶の性質を明らかにする．本章ではパラメータを減らし解析を容易にするため，$w_1 = w_2 = \cdots = w_{n-1} (= w)$ とする．

5.1 全転送コスト

各レベルの転送コストを統一的に扱うため，レベル i, $i+1$ 間の転送コストは i 次フェッチ数に定数 $C_{i,i+1}$ をかけたものとする．このとき，前章の n 階層記憶モデルの全転送コスト T_n は，

$$T_n = \sum_{i=1}^{n-1} C_{i,i+1} F_i$$

$$= \sum_{i=1}^{n-1} C_{i,i+1} q_{i+1} \log_w q_{i+1} \tag{13}$$

となる．ただし，F_i としては，式(12)の最悪値を用いた．

1つの i 次ページは $\prod_{j=1}^{i} p_j$ 個の0次ページ（レコード）からなるので，

$$C_{i,i+1} = A_{i,i+1} + B_{i,i+1} \prod_{j=1}^{i} p_j \tag{14}$$

と表す．ここで，$A_{i,i+1}$ はレベル i, $i+1$ 間のアクセス時間，$B_{i,i+1}$ はレベル i, $i+1$ 間の1レコードあたりの転送時間である．

5.2 4階層，2階層，3階層記憶の比較

4階層，2階層，3階層記憶のモデルを図8に示す．各モデルとも最上位と最下位の記憶装置は同じものとする．したがって，最下位レベルとその上のレベルとの転送時間パラメータである A_{cd}, B_{cd} は共通である．

本節では，次の事項について調べる．

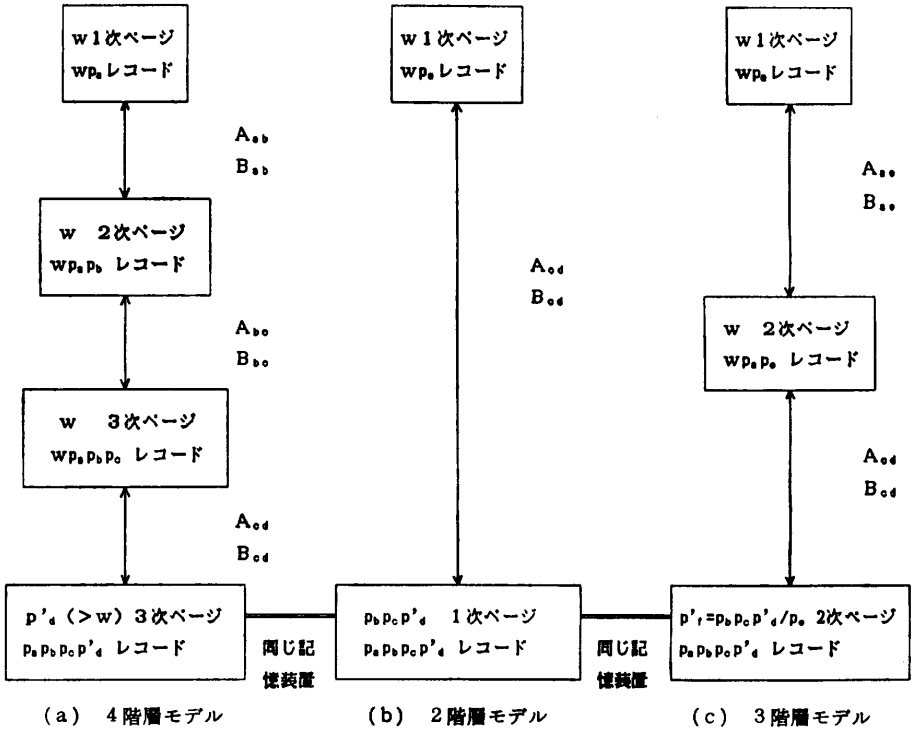

図8 記憶階層モデル
Fig. 8 Models of memory hierarchy.

1) 3階層と4階層の場合，中間レベルの記憶容量はどの程度あれば十分か．

2) 4階層と2階層では，どの程度転送コストがちがうか．

3) 3階層で4階層とほぼ同じ効率を得るためには，3階層の中間レベルにどのような記憶をおいたらよいか．

図8中の転送パラメータ A_{ab}, A_{bc}, A_{cd}, A_{ae}, B_{ab}, B_{bc}, B_{bd}, B_{ae} を用いると，式(11)，(13)，(14)から，レコード並べかえに必要な全転送コストの最大値は，それぞれの場合について，

$$T_4=(A_{ab}+B_{ab}p_a)p_bp_cp_d'\log_w p_bp_cp_d' + (A_{bc}+B_{bc}p_ap_b)p_cp_d'\log_w p_cp_d' + (A_{cd}+B_{cd}p_ap_bp_c)p_d'\log_w p_d', \quad (15)$$

$$T_2=(A_{cd}+B_{cd}p_a)p_bp_cp_d'\log_w p_bp_cp_d', \quad (16)$$

$$T_3=(A_{ae}+B_{ae}p_a)p_bp_cp_d'\log_w p_bp_cp_d' + (A_{cd}+B_{cd}p_ap_e)\frac{p_bp_cp_d'}{p_e}\log_w\frac{p_bp_cp_d'}{p_e} \quad (17)$$

となる．式(15)の第1，2，3項はそれぞれレベル1，2間，2，3間，3，4間の転送コストである．式(16)，(17)についても同様である．

比較を容易にするため，仮定を設け近似を行う．1フェッチあたりのアクセス時間とデータ転送時間比は，転送効率と，メモリ使用率間のトレードオフから，適当な値が存在する．4階層モデルにおいては，この比が各レベル間とも一定で，$1:\alpha$ になるよう転送単位が調整されていると仮定する．すなわち，

$$\frac{B_{ab}p_a}{A_{ab}}=\frac{B_{bc}p_ap_b}{A_{bc}}=\frac{B_{cd}p_ap_bp_c}{A_{cd}}=\alpha \quad (18)$$

である．さらに，やや粗い近似ではあるが，式(15)において，対数 ($\log_w$) がかかっている値はそうでない値にくらべて変化がゆるやかなので，ほぼ同じであるとみなす．すなわち，

$$\log_w p_bp_cp_d'\fallingdotseq\log_w p_cp_d'\fallingdotseq\log_w p_d' \quad (19)$$

とする．このとき，式(15)から T_4 は，

$$T_4\fallingdotseq(1+\alpha)(A_{ab}p_bp_c+A_{bc}p_c+A_{cd})p_d'\log_w p_bp_cp_d' \quad (15)'$$

となる．特定のレベル間がボトルネックにならないようにするため，式(15)′の2番目のかっこ中の項を等しくする．

$$A_{ab}p_bp_c=A_{bc}p_c=A_{cd}. \quad (20)$$

式(18)，(20)から，

$$B_{ab}=B_{bc}=B_{cd}. \quad (21)$$

すなわち，1レコードあたりのデータ転送時間は，各記憶階層間でほぼ同じでなければならないことがわかる．また，式(20)から，

$$p_b=\frac{A_{bc}}{A_{ab}},\ p_c=\frac{A_{cd}}{A_{bc}} \quad (22)$$

が求まる．このとき，

$$T_4\fallingdotseq3(1+\alpha)A_{ab}p_bp_cp_d'\log_w p_bp_cp_d' \quad (15)''$$

である．一方，式(18)から，$B_{cd}p_a=\alpha A_{cd}/p_bp_c$ でこの値は A_{cd} にくらべずっと小さいと考えてよいので式(16)は，

$$T_2\fallingdotseq A_{cd}p_bp_cp_d'\log_w p_bp_cp_d' \quad (16)'$$

と書ける．式(15)″と(16)′から，

$$\frac{T_2}{T_4}\fallingdotseq\frac{A_{cd}}{3(1+\alpha)A_{ab}} \quad (23)$$

となる．

一方，3階層モデルについても，4階層モデルと同様，

$$\frac{B_{ae}p_a}{A_{ae}}=\frac{B_{cd}p_ap_e}{A_{cd}}=\alpha \quad (18)'$$

$$\log_w p_bp_cp_d'\fallingdotseq\log_w(p_bp_cp_d'/p_e) \quad (19)'$$

が言えるとし，式(17)から，

$$T_3\fallingdotseq(1+\alpha)(A_{ae}p_bp_c+A_{cd}p_bp_c/p_e)p_d'\log_w p_bp_cp_d' \quad (17)'$$

を得る．特定のレベル間がボトルネックにならないようにするため，

$$A_{ae}=A_{cd}/p_e \quad (24)$$

とする．このとき，

$$T_3\fallingdotseq2(1+\alpha)A_{ae}p_bp_cp_d'\log_w p_bp_cp_d' \quad (17)''$$

である．$T_3\fallingdotseq T_4$ とおくと，

$$A_{ae}\fallingdotseq(3/2)A_{ab}. \quad (25)$$

式(24)から，$p_e=A_{cd}/A_{ae}$ であるが，これに式(25)，(22)を代入すると，

$$p_e\fallingdotseq(2/3)p_bp_c \quad (26)$$

が得られる．

以上から，結果をまとめると，

1) 中間レベルの記憶容量は，4階層モデルのとき，レベル2はレベル1の $p_b=A_{bc}/A_{ab}$ 倍，レベル3はレベル2の $p_c=A_{cd}/A_{bc}$ 倍，3階層モデルのとき，レベル2はレベル1の $p_e=A_{cd}/A_{ae}$ 倍，それぞれあれば十分である．これらの値はアクセス時間比になっている．

2) 2階層，4階層モデルの全転送コストの比は，

$$\frac{T_2}{T_4}\fallingdotseq\frac{A_{cd}}{3(1+\alpha)A_{ab}}$$

である．

3） 3階層モデルで，4階層モデルと同程度の効率を得るために必要な3階層の中間レベルの記憶装置は，

アクセス時間：$A_{ac}=(3/2)A_{ab}$,

容量：$wp_ap_c=(2/3)wp_ap_bp_c$ レコード．

すなわち，4階層モデルでのレベル2のアクセス時間×3/2で，レベル3の容量の2/3をもつものでなければならない．

現実の記憶装置にあてはめてみる．4つのレベルとして，主記憶 (MOS-DRAM)，半導体ディスク，磁気バブル，磁気ディスクを想定する．それぞれのアクセス時間，データ転送レートの典型例[4),5)]と1レコードを100Bとしたときの1レコードあたりのデータ転送時間を**表1**に示す．この表から $p_b=A_{bc}/A_{ab}\fallingdotseq 17$ だからレベル2はレベル1の17倍，$p_c=A_{cd}/A_{bc}=5$ だからレベル3はレベル2の5倍の記憶容量があればよい．式(21)の条件から，データ転送時間は磁気バブルがボトルネックになっていることがわかる．したがって，レベル2，4のデータ転送レートの高速性が活されず，実質的なデータ転送時間は B_{bc} と同程度に制限される．$p_a=3$（1次ページが $100\times3=300$ B）とすると，式(18)から $\alpha=1.25$ となる．このとき式(23)から，

$$T_2/T_4\fallingdotseq 12 \tag{27}$$

となる．

より具体的に，磁気ディスク上の250,000個のレコードを並べかえる場合についてコストを計算してみる．1レコード100Bだから総データ量は約250MBで $p_d'=10,000$ になる．$w=1,000$ とすると，レベル1，2，3の必要な記憶容量はそれぞれ300kB，5MB，25MB，各レベル間の転送単位はレベル1-2間，2-3間，3-4間の順に300B，5kB，25kBである．式(15)，(16)から，T_4，T_2 を計算すると，

$$T_4\fallingdotseq 2,500\text{秒}，\quad T_2\fallingdotseq 43,000\text{秒} \tag{28}$$

となる．2章で述べたように書出しも読込みと同じデータ転送回数を要するので並べかえにかかる時間を見積るためには2倍する必要がある．式(28)から，$T_2/T_4\fallingdotseq 17$ となる．式(27)と差が生じているのは，式(15)，(16)から式(23)を得るのにいくつかの近似を行ったためである．

5.3 n 階層記憶の場合

比較を容易にするため，5.2節後半と同様な仮定と近似を行う．すなわち，

$$\frac{B_{12}p_1}{A_{12}}=\frac{B_{23}p_1p_2}{A_{23}}=\cdots=\frac{B_{n-1,n}p_1p_2\cdots p_{n-1}}{A_{n-1,n}}=\alpha, \tag{29}$$

$$\log_w p_2p_3\cdots p_{n-1}p_n'\fallingdotseq\log_w p_3p_4\cdots p_{n-1}p_n'\fallingdotseq\cdots\fallingdotseq\log_w p_n' \tag{30}$$

を仮定すると，

$$T_n\fallingdotseq(1+\alpha)(A_{12}p_2p_3\cdots p_{n-1}+A_{23}p_3p_4\cdots p_{n-1}+\cdots+A_{n-1,n})p_n'\log_w p_2p_3\cdots p_{n-1}p_n' \tag{31}$$

となる．上式の後ろのかっこ内の項を等しくすると，

$$A_{12}p_2p_3\cdots p_{n-1}=A_{23}p_3p_4\cdots p_{n-1}=\cdots=A_{n-1,n}$$

$$\therefore\quad p_2=\frac{A_{23}}{A_{12}},\ p_3=\frac{A_{34}}{A_{23}},\cdots,\ p_i=\frac{A_{i,i+1}}{A_{i-1,i}},\cdots,$$

$$p_{n-1}=\frac{A_{n-1,n}}{A_{n-2,n-1}} \tag{32}$$

が求まる．このとき，

$$T_n\fallingdotseq(n-1)(1+\alpha)A_{12}p_2p_3\cdots p_{n-1}p_n'\times\log_w p_2p_3\cdots p_{n-1}p_n' \tag{31}'$$

である．一方，T_2 は $n\geqq 3$ のとき，

$$T_2\fallingdotseq A_{n-1,n}p_2p_3\cdots p_{n-1}p_n'\log_w p_2p_3\cdots p_{n-1}p_n' \tag{33}$$

と書けるので，

$$\frac{T_2}{T_n}\fallingdotseq\frac{A_{n-1,n}}{(n-1)(1+\alpha)A_{12}} \tag{34}$$

である．アクセス時間の逆数はアクセススピードになるので，n 階層モデルのほうが2階層モデルにくらべ“レベル2のアクセススピード／最下位レベルのアクセススピード”に比例して高速に並べかえできることがわかる．

さらに，n 階層モデルのレベル i と $i+1$ ($i=2,3,\cdots,n-2$) の2つのレベルのかわりに1つのレベルで同程度の効率を得る方法を考える．このレベルを $i+0.5$ とする．**図9**に転送パラメータとともに示す．

このとき，

$$\frac{B_{i-1,i+0.5}p_1p_2\cdots p_{i-1}}{A_{i-1,i+0.5}}=\frac{B_{i+1,i+2}p_1p_2\cdots p_{i-1}p_{i+0.5}}{A_{i+1,i+2}}=\alpha, \tag{35}$$

表1 記憶装置の特性
Table 1 Characteristics of storages.

レベル	記憶装置	アクセス時間	データ転送レート	データ転送時間
1	MOS-DRAM	100 ns	—	—
2	半導体ディスク	0.3 ms (A_{ab})	3 MB/s	33 μs/100 B (B_{ab})
3	磁気バブル	5 ms (A_{bc})	0.8 MB/s	125 μs/100 B (B_{bc})
4	磁気ディスク	25 ms (A_{cd})	3 MB/s	33 μs/100 B (B_{cd})

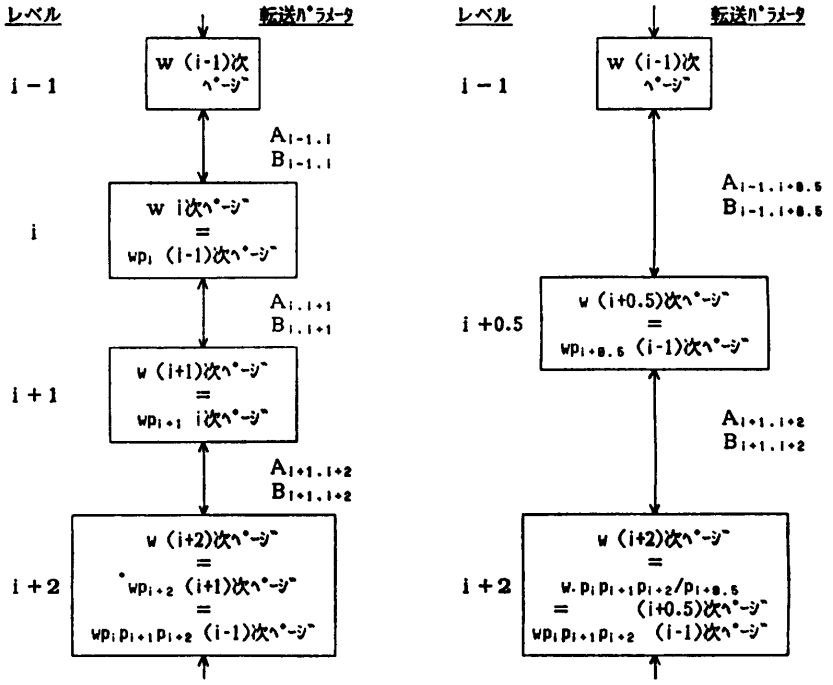

図 9 n 階層モデルから $n-1$ 階層モデルへ
Fig. 9 Models of n-level and $(n-1)$-level storage.

$$\log_w p_i p_{i+1}\cdots p_{n-1}p_n' \fallingdotseq \log_w \frac{p_i p_{i+1}\cdots p_{n-1}p_n'}{p_{i+0.5}} \tag{36}$$

を仮定すると，全転送コスト T_{n-1} は，

$$\begin{aligned} T_{n-1} \fallingdotseq (1+\alpha)(&A_{12}p_2p_3\cdots p_{n-1} + A_{23}p_3p_4\cdots p_{n-1} + \cdots \\ &+ A_{i-1,i+0.5}p_i p_{i+1}\cdots p_{n-1} \\ &+ A_{i+1,i+2}\frac{p_i p_{i+1}\cdots p_{n-1}}{p_{i+0.5}} \\ &+ A_{i+2,i+3}p_{i+3}p_{i+4}\cdots p_{n-1} + \cdots + A_{n-1,n})p_n' \\ &\times \log_w p_2p_3\cdots p_{n-1}p_n' \end{aligned} \tag{37}$$

となる．特定のレベルがネックにならない条件から，

$$p_{i+0.5} = A_{i+1,i+2}/A_{i-1,i+0.5} \tag{38}$$

が求まる．このとき，

$$\begin{aligned} T_{n-1} \fallingdotseq (n-2)(1+\alpha)&A_{i-1,i+0.5}p_i p_{i+1}\cdots p_{n-1}p_n' \\ &\times \log_w p_2p_3\cdots p_{n-1}p_n' \end{aligned} \tag{37}'$$

となる．式(31)′は(29)から

$$\begin{aligned} T_n \fallingdotseq (n-1)(1+\alpha)&A_{i-1,i}p_i p_{i+1}\cdots p_{n-1}p_n' \\ &\times \log_w p_2p_3\cdots p_{n-1}p_n' \end{aligned} \tag{31}''$$

と書くこともできる．$T_{n-1} \fallingdotseq T_n$ とおくと，

$$A_{i-1,i+0.5} = \frac{n-1}{n-2}A_{i-1,i} \tag{39}$$

となる．式(38)に(39)，(32)を代入すると，

$$p_{i+0.5} = \frac{n-2}{n-1}p_i p_{i+1} \tag{40}$$

が得られる．

以上の計算から次のことが言える．

1）中間レベルの記憶容量に対して，式(32)から $i=1,2,\cdots,n-2$ についてレベル $i+1$ はレベル i の $p_i = A_{i,i+1}/A_{i-1,i}$ 倍あれば十分である．

2）最下位レベルの記憶装置が同じとき，よく調整された n (≧3) 階層モデルと2階層モデルの全転送コストの比は，

$$\frac{T_2}{T_n} \fallingdotseq \frac{A_{n-1,n}}{(n-1)(1+\alpha)A_{12}}$$

である．

3）レベル i, $i+1$ ($i=2,3,\cdots,n-2$) を1つのレベル $i+0.5$ におきかえて，しかも同じ効率を得るためには，

アクセス時間：$A_{i-1,i+0.5} = \dfrac{n-1}{n-2}A_{i-1,i}$

容量：$\dfrac{n-2}{n-1}w$ ($i+1$) 次ページ

すなわち，レベル i のアクセス時間×$(n-1)/(n-2)$ で，レベル $i+1$ の $(n-2)/(n-1)$ の容量をもたなけれ

ばならない．この結果は中間レベルの記憶階層数を減らす場合の指標となる．

6. むすび

多階層記憶におけるレコード並べかえについて論じた．提案の並べかえアルゴリズム（拡張 $d\log_w d$ 法）を用いると，レベル $i, i+1$ 間については，レベル間の1回のデータ転送量（i 次ページ）を単位とし，レベル i の記憶容量を w_i とすると，たかだか $q_{i+1}\log_{w_i} q_{i+1}$ フェッチで並べかえできることがわかった．ここで，q_{i+1} は並べかえるべきデータ量を i 次ページで表現したものである．また，全転送コストの定義を行い，1）中間レベルに必要十分な記憶容量，2）よく調整された n 階層モデルと2階層モデルの全転送コストの比，3）中間レベルの記憶階層数を1つ減らした場合，もとの記憶システムと同等な効率を得るための記憶，について解析を行った．

参考文献

1) Tsuda, T. and Sato, T.: Transposition of Large Tabular Data Structures with Applications to Physical Database Organization (Part 1) Transposition of Tabular Data Structures, *Acta Inf.*, Vol. 19, No. 1, pp. 13-33 (1983).
2) Tsuda, T., Sato, T. and Tatsumi, T.: Minimizing Page Fetches for Permuting Information in Two-Level Storage (Part 1) Generalization of the Floyd Model, *J. Inf. Proc.*, Vol. 6, No. 2, pp. 74-77 (1983).
3) 佐藤，津田：2階層記憶における効率のよいデータ並べかえアルゴリズム，情報処理学会論文誌，Vol. 27, No. 9, pp. 845-852 (1986).
4) 亀山ほか：磁性体メモリ，情報処理，Vol. 27, No. 6, pp. 618-629 (1986).
5) 中村：記憶装置が多様化，急がれる記憶階層マネジメントの実施，日経コンピュータ，No. 149, pp. 71-84 (1987).

（昭和62年2月18日受付）
（昭和63年2月10日採録）

佐藤 隆士（正会員）

昭和28年9月生．昭和53年岡山大学大学院修士課程（電子工学専攻）修了．工学博士．同年詫間電波高専助手．現在，同校助教授．データベース，記憶階層の研究に従事．電子情報通信学会，CAI 学会各会員．

津田 孝夫（正会員）

1932年生．1957年京都大学工学部電気工学科卒業．現職は京都大学工学部情報工学科教授．工学博士．現在の主要研究テーマは，メモリ階層間データ転送量の下界とそれによるアルゴリズムの最適化，ベクトル計算機のための自動ベクトル化と自動並列化，実時間オペレーティングシステムなど専用 OS の構成と実現法など．